Game report data prediction model based on LSTM-GRU

Jiayi Wang [#], Zhi Liu [#, *], Jiaqing Li [#]

Department of computer science and technology, Yunnan university of finance and economic, Yunnan, China, 650399

* Corresponding Author Email: LZ1840515884@163.com

[#]These authors contributed equally.

Abstract. With the increasing number of participants in wordles game, the number of reports changed, we firstly preprocessed the data by sliding window, and then analyzed the reasons for the change of the number of reports and predicted them by LSTM+GRU time series prediction model, and got that the change of the number of reports increased slowly in the early stage and decreased gradually in the later stage, then we predicted the change of the number of reports on March 1, and its change The interval is [22233, 22407], and it has a good fitting effect. Finally, through the comparison of different models, this model has the smallest error, which increases the credibility of the model. Finally, it is found that if this game does not innovate in time, the number of people will gradually decrease in the later.

Keywords: LSTM-GRU; Prediction Model; Machine learning; Game repot data.

1. Introduction

Wordle is a popular puzzle game currently offered daily by The New York Times on social media platforms, in which players have a total of six chances to guess a word consisting of five letters. Each time a guess is made, a hint is given based on how well the guessed word matches the correct answer. The game play in hard mode is different from the easy mode and is more difficult. Currently wordle is supported in more than 60 languages worldwide. The number of reports is changing daily as the number of players increases.

Based on the changes in the number of reports we want to predict the changes and explain their causes by models.[1] Conventional methods include time series prediction[2],artificial neural network prediction, etc.The mathematical theory of the above methods is complete, but the results in practice are not satisfactory, difficult and error-prone, and usually the prediction time interval is long, the reliability of the results predicted using the above methods will be significantly reduced.

The problem of dynamic temporal prediction is solved to some extent using recurrent neural networks (RNN) [3], but changes in player reported scores occur daily, and to further understand the connection between historical and current reported scores, RNNs are required to memorize historical and current data. To this end, an improved dynamic long-short memory model (LSTM-GRU) is cited in this paper to improve the memory capability of RNNs and test their feasibility by fitting them with various machine learning methods, so as to improve the accuracy of the prediction model and be able to reasonably explain current changes, or be able to provide theoretical basis and methodological suggestions for games or platforms similar to it [4].

2. Model

2.1. Neural Network Model（RNN）

RNNs are used in a variety of fields, including natural language processing , machine translation, speech recognition, image processing, text similarity processing , and recommendation systems, etc. They are capable of extracting feature data from the in-put data. The latter is used for visual processing such as image [5]. The RNN network model is shown in figure 1

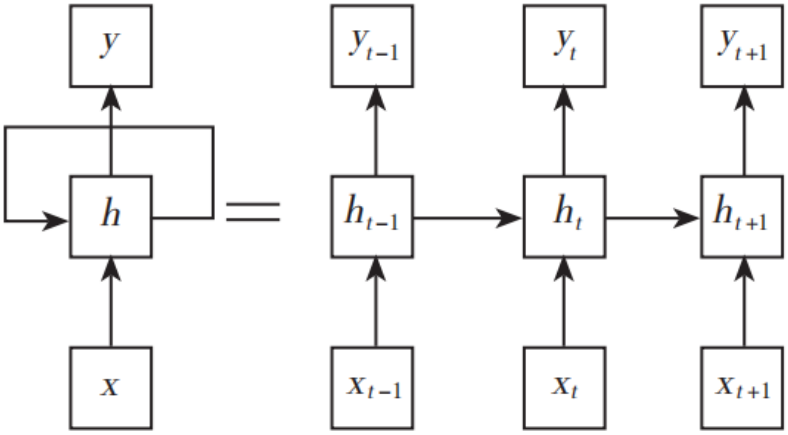


Figure 1. RNN network structure.

2.2. LSTM-GRU

LSTM (Long short-term memory) is a special kind of RNN that solves the problem of predictability of long time series due to the lag of unknown duration that may exist between important events in the time series, and the relative insensitivity to the gap length is a major advantage of LSTM [6]. the LSTM unit consists of an input gate, a forgetting gate, an output gate and a unit state, which remembers the values in any time interval, and three gates control the flow of information in and out of the cell.

$$I_t=\sigma\left(X_tW_{xi}+H_{t-1}W_{hj}+b_i\right) \tag{1}$$

$$F_t=\sigma(X_tW_{xf}+H_{t-1}W_{hf}+b_f) \tag{2}$$

$$O_t=\sigma(X_tW_{xo}+H_{t-1}W_{ho}+b_o) \tag{3}$$

$$\widehat{C_{t1}}=\tan h(X_tW_{xc}+H_{t-1}W_{hc}+b_c) \tag{4}$$

$$C_t=\sigma\left(X_tW_{xf}+H_{t-1}W_{hf1}+b_f\right) \tag{5}$$

$$H_t=O_t\odot\tan h(C_t) \tag{6}$$

GRU, as a variant of LSTM [7], synthesizes the forget gate and the input gate into a single update gate. It also mixes cell states and hidden states, plus some other modifications [8]. With almost no difference in prediction accuracy, the algorithm and model of GRU [9] are more compact compared to LSTM because GRU reduces the training parameters to obtain higher convergence speed. The GRU network model is shown in figure 2

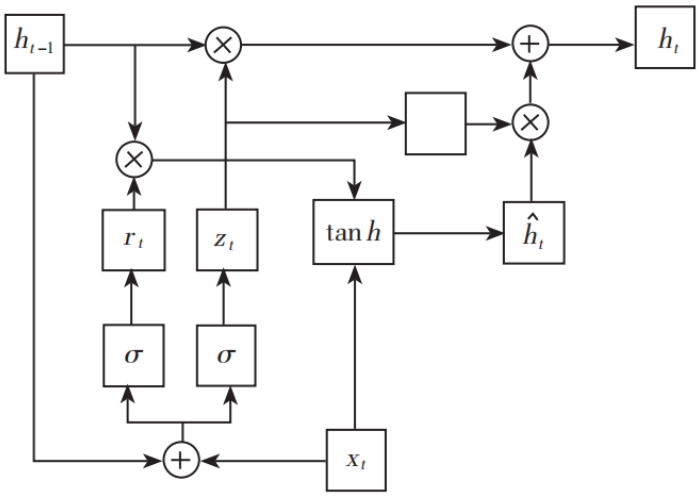


Figure 2. GRU network structure.

$$R_t=\sigma(X_t W_{xr}+H_{t-1} W_{hr}+b_r) \quad (7)$$

$$Z_t=\sigma(X_t W_{xz}+H_{t-1} W_{hz}+b_z) \quad (8)$$

$$\widehat{H_t}=\tan h(X_t W_{xh})+R_t \odot H_{t-1} W_{hh}+b \quad (9)$$

$$H_t=Z_t \odot H_t+(1-Z_t) \odot H_t \quad (10)$$

2.3. Data pre-processing

There are many factors that affect the scores reported on the same day, and using daily data will result in a less efficient model with too many operations when only internal factors of the scores are considered, so we use pandas to perform a sliding window transformation for a group of 10 days and find its mean value for calculation and prediction [10].The results of data processing are shown in the table 1.

Table 1. The results of data processing.

Date	Word	rolling10
2022/1/18	proxy	165092.3
2022/1/28	perky	277533.3
2022/2/7	elder	329959.1

3. Results

3.1. Analysis of experimental results

After the LSTM-GRU model prediction, we found that the trend of change is increasing in the first period and gradually decreasing after reaching a certain peak. The data reported with this model for March 1, 2023 is: 23320, and the model error is: 0.0039 (0.39%).

We fit [11] the results obtained from 100 iterations of training and found that the MAPE value is 7.55, which is a good fit, as shown in figure 3.

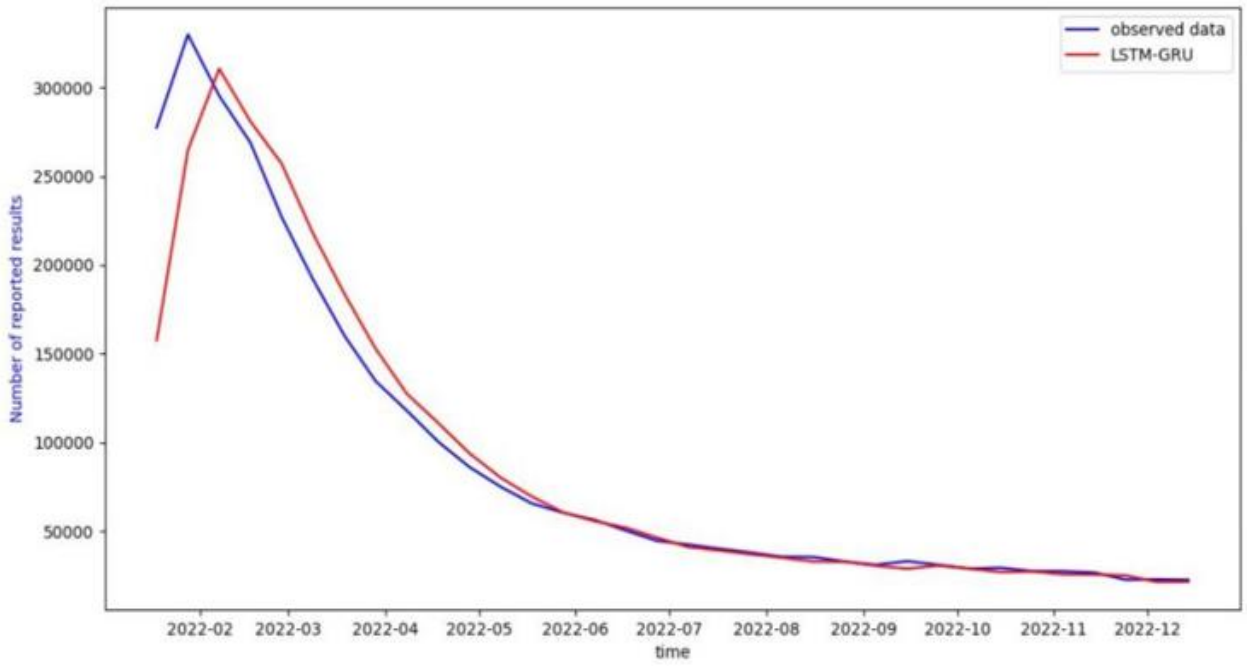


Figure 3. LSTM-GRU Fitting results.

From the above discussion, it can be seen that for dynamic time prediction and prediction over a large time span, the use of the LSTM-GRU method results in a small error and a high degree of fit, indicating that for similar dynamic time prediction problems in game classes or social platforms, the

LSTM-GRU model has a greater adaptability and practical ability and higher prediction accuracy compared with traditional time prediction models. Therefore, we conclude that: the reported data obtained by using GRU+LSTM model prediction has higher credibility. Combining their model errors, we can obtain the prediction interval for the reported data on March 1, 2023 [22233, 22407].

4. Discussion

In order to fully validate or avoid possible errors in the above models, we used linear regression, decision tree, and random forest models to forecast the March 1 reported data, and the relevant results are shown in table 2:

Table 2. Other model prediction data and errors.

Model	Forecast data	Error value
Linear regression	23483	0.0070
Decision tree	25442	0.0912
Random forest	23879	0.0232
GRU	23320	0.0039

The comparison shows that the LSTM-GRU model has a lower error, further confirming the accuracy of our model.

For the LSTM-GRU model, we first performed data preprocessing, removed outliers and constructed a sliding window transformation. Secondly, we built the prediction model for the number of reports and after the implementation of different algorithms with layers of comparison have relatively small errors, and improve the credibility of the model. Finally, the predicted value is basically consistent with the trend of change of the true value, the data tracking effect is better, the prediction accuracy is high, and the prediction accuracy on the number of reported results is higher and better than that using a single prediction model. However, our model does not take into account the influence of external factors, which can cause a certain amount of error.

5. Conclusion

Games are a product of the Internet era, and they will accompany us for a long time, both as customers and developers. And by analyzing the information behind them, we are able to get useful data to update the games and attract players. The model is a typical model to predict the reported number of participants behind the game, and our model LSTM-GRU is more accurate and practical than other prediction models.

References

[1] WANG Guifeng. Research on trend prediction modeling method based on situational information [D].University of Science and Technology of China,2021. DOI:10.27517/d.cnki.gzkju.2021.000484.)

[2] ZHOU Sunyu, JIN Baoxuan, ZHAO Kang, et al. Research on high-performance intersection method based on time prediction model[J].Geography and Geo-Information Science,2020,36(2):8-15.)

[3] SUN Yuhang,LIU Yang. Prediction of shear wave velocity using GRU neural network[J].Oil Geophysical Prospecting,2020,55(3):484-492.)

[4] Althelaya K A, El-Alfy E s M, Mohammed S. Stock Market Forecasting Using Bidirectional and Stacked Multivariate Analysis (LSTM, GRU) [C]//2018 21st Saudi Computer Society National Computer Conference (NCC). IEEE,2018:1-7.

[5] YANG Fan. Research on short-term wind power prediction technology based on deep learning[D].Huazhong University of Science and Technology, 2020.)

[6] Rodrigue A K.Research on prediction method based on LSTM[D].Anhui Agricultural University,2019.

[7] ZHOU Zhangyuan,HE Xiaoling. Stock price prediction method based on optimized LSTM model [J/OL].Statistics and Decision,2023(06):143-148[2023-03-26]. DOI:10.13546/j.cnki.tjyjc.2023.06.026.

[8] CAI Hui,LUO Jiawei. Comparative analysis of ICU patient data on the effect and convergence rate of LSTM and GRU prediction models[J].Journal of Mianyang Normal University,2020.)

[9] GUO Hongcun,ZANG Zhaoxiang,ZHANG Zhen,TIAN Pei. Prediction of instruction types in real-time strategy games by GRU improved algorithm[J].computer programing skills and maintenance,2021(08):3-6.DOI:10. 16184/j.cnki.comprg.2021.0 8.001.)

[10] LIU Tianpeng, NIU Feng, ZHANG Shiyong, et al. Radar signal sorting based on timing matching and variable-length sliding window[J].Radar Science and Technology,2020,18(5):562-567,578.)

[11] ZHANG Lin. Fitting and prediction of general growth model of new crown pneumonia epidemic spread[J].Journal of University of Electronic Science and Technology of China, 2020,49(3):345-348.)